A proposed methodology for the assessment of arsenic, nickel, cadmium and lead levels in ambient air

Germán Santos[1*], Ignacio Fernández-Olmo[1]

[1]Dpto. de Ingenierías Química y Biomolecular, Universidad de Cantabria, Santander, 39005, Spain

[*] Corresponding author. Tel.: +34 942201579; fax: +34 942201591

E-mail address: santosg@unican.es

Abstract: Air quality assessment, required by the European Union (EU) Air Quality Directive, Directive 2008/50/EC, is part of the functions attributed to Environmental Management authorities. Based on the cost and time consumption associated with the experimental works required for the air quality assessment in relation to the EU-regulated metal and metalloids, other methods such as modeling or objective estimation arise as competitive alternatives when, in accordance with the Air Quality Directive, the levels of pollutants permit their use at a specific location. This work investigates the possibility of using statistical models based on Partial Least Squares Regression (PLSR) and Artificial Neural Networks (ANNs) to estimate the levels of arsenic (As), cadmium (Cd), nickel (Ni) and lead (Pb) in ambient air and their application for policy purposes. A methodology comprising the main steps that should be taken into consideration to prepare the input database, develop the model and evaluate their performance is proposed and applied to a case of study in Santander (Spain). It was observed that even though these approaches present some difficulties in estimating the individual sample concentrations, having an equivalent performance they can be considered valid for the estimation of the mean values –those to be compared with the limit/target values– fulfilling the uncertainty requirements in the context of the Air Quality Directive. Additionally, the influence of the consideration of input variables related to atmospheric stability on the performance of the studied statistical models has been determined. Although the consideration of these variables as additional inputs had no effect on As and Cd models, they did yield an improvement for Pb and Ni, especially with regard to ANN models.

Keywords: air quality assessment; atmospheric stability; methodology; metals; statistical models

1. Introduction

The member states of the European Union are obliged by the Air Quality Framework Directive, Directive 2008/50/EC (EC, 2008), to carry out the air quality assessment of certain specified pollutants according to a group of common methods and criteria. These methods vary depending on the population within each air quality zone or agglomeration and the relative levels of the mentioned pollutants compared with the respective assessment thresholds, which are expressed as a percentage of the corresponding limit/target value (EC, 2004; EC, 2008). Thus, fixed measurements are required when it has been confirmed that the levels of an atmospheric pollutant exceed the upper assessment threshold. A combination of fixed measurements and modeling techniques and/or indicative measurements is possible when the pollutant concentration is between both assessment thresholds. Moreover, air quality assessment based exclusively on modeling and/or objective estimation is solely reserved for zones and agglomerations with relatively good air quality and no large conurbations, in which at a preliminary assessment stage the levels of those pollutants whose air quality is to be assessed consistently fall sufficiently below their respective lower assessment thresholds (EC, 2008). In this sense, the minimum quality requirements for the objective estimation techniques to be acceptable in the context of Directive 2008/50/EC, which include uncertainty and minimal data capture and time coverage, are less strict than those for modelling techniques and experimental measurements, either fixed or indicative. Specifically, the relative uncertainty should be lower than 100 % (EC, 2008). Although no indications on the nature of these methods are provided in the European Directives, the European Commission in its report for guidance on assessment under the European Union Air Quality Directives interprets objective estimation techniques as "mathematical methods to calculate concentrations from values measured at other locations and/or times, based on scientific knowledge of the concentration distribution". In this context, statistical models can be regarded as objective estimation techniques in the sense that they are based on statistical data analysis establishing empirical relationships between ambient concentrations and meteorological variables instead of simulating the relationship between emissions and immission concentrations through the description of the physical phenomena that rules the air transportation and dispersion of pollutants in the atmosphere, as with classical deterministic models (Daly and Zanetti, 2007). It consists of generating estimations of the European regulated pollutant concentrations directly from regional air quality information, which indirectly reflects atmospheric pollution in part due to local emission sources. Therefore, there is an interest in deepening the understanding of this sort of methods to be used as air quality assessment tools

as they are cost-effective and simpler than other approaches, such as air dispersion modelling which can also be considered more affordable than experimental measurements, especially for certain pollutants, namely metals and polycyclic aromatic hydrocarbons, whose analytical determination, unlike other atmospheric pollutants, is expensive and time consuming.

The use of modelling techniques is encouraged even when continuous monitoring is compulsory because of its capacity of providing supplementary information to air quality monitoring and future projections regarding public exposure under different emission scenarios. Furthermore, modelling is becoming a prominent air quality assessment tool as it has been reported that different modelling approaches have already been incorporated as part of the routine air quality network monitoring and assessment procedures in a number of countries, e.g., Finland, Norway, Sweden or United Kingdom. For those reasons, the use of models and other alternative air quality assessment tools for policy support is expected to continue increasing in the future.

Although the use of statistical modelling techniques to predict immission concentrations of major atmospheric pollutants has been extensively investigated in the literature, there are not many studies addressing the estimation of the levels of compounds bound to particulate matter; metals in particular. Artificial neural networks (ANNs), state-space modelling, time series autoregressive modelling and linear regression stand out among the approaches used to that end. With respect to ANNs, Chelani et al. (2002) estimated the ambient air levels of Cd, Cr, Fe, Ni, Pb, Zn as well as PM_{10} in the city of Jaipur, India, by means of artificial neural network models obtaining low values of root mean square error (RMSE). In addition, Li et al. (2009) reconstructed occupational manganese exposure by means of back-propagation artificial neural networks and multiple linear regression. In a previous study, a state-space model coupled with Kalman filter and an autoregressive model with external input (ARX model) was used by Chelani et al. (2001) to predict the concentration of Pb, Fe and Zn and respirable suspended particulate matter in Delhi. Apart from ANNs, some research has been conducted to model metal concentrations in ambient air using other statistical approaches. Hernández et al. (1992) applied state-space modelling, Box-Jenkins modelling and time series autoregressive integrated moving average (ARIMA) models to estimate the daily concentrations of air-particulate Fe and Pb in Madrid (Spain). Predictions of daily Fe were better than those of Pb. No difference being found between State-space and Box-Jenkins models, their outcomes were better than those of ARIMA models in terms of root mean squared error (RMSE), correlation coefficient and efficiency. The state space model performed better than the ARX model. On the other

hand, Vicente et al. (2012) developed predictive models based on multiple regression analysis together with time series (ARIMA) models to predict the concentration of total suspended particles (TSP), PM_{10}, As, Cd, Ni and Pb in the ambient air of Castellón (Spain). Arruti et al. (2011) developed statistical models based on multiple linear regression and principal component regression as objective estimation techniques to estimate the immission levels of As, Cd, Ni and Pb at four sampling sites in Cantabria (Northern Spain). Both techniques were found to be valid approaches; however, there was room for improvement with regard to their performance.

The implications of the use of objective estimation techniques by local air quality management authorities to assess air quality when it is possible are discussed in this work. For this purpose, we present a case of study focused on the estimation of ambient air levels of Pb, As, Ni and Cd in Santander (Spain), by means of statistical models based on partial least squares regression (PLSR) and artificial neural networks.

Additionally, given that normally deterministic models take into consideration directly or indirectly the atmospheric stability to simulate the dispersion of pollutants in the atmosphere, this work also aims at determining the influence of the use of input meteorological variables related to atmospheric stability on the performance of the studied statistical models.

2. Materials and methods

2.1. Case of study: European Union regulated metal(loids) in Santander, Spain

Santander, with approximately 176000 inhabitants in 2014 (INE, 2014), is the capital of the Cantabria Region in the North of Spain. The city is extended over a bay and in its suburbs, at 5-10 km, there is an industrial area mainly related to steel and ferroalloys manufacturing plants that makes Santander as the most complex urban area in the region. A sampling campaign was conducted by our research group to determine daily values of PM_{10} (particulate matter with and aerodynamic diameter lower than 10 μm) in 2008, 2009 and 2011 in a coastal sampling site (ETSIIT) located in the rooftop of the building "Escuela Superior de Ingenieros Industriales y de Telecomunicación" (43º 28' 24" N, 3º 47' 54" W at 23 m.a.s.l) and approximately 1 km far from the sea. Glass microfiber filters (Whatman, 150 mm diameter) were used for PM_{10} sampling during 2008 and 2009, whereas the 2011 campaign was carried out with quartz microfiber filters (Sartorius AG, 47 mm diameter) because of their lower detection limits, mainly for As (0.03 vs. 0.54 ng/m^3) and Cd (0.01 vs. 0.11 ng/m^3). Metal levels of 2008 and 2009 are reported in Arruti et al. (2011). In that work, the content of Pb, As, Cd, Ni and other trace metals on PM_{10} was determined

by microwave assisted acidic digestion prior to ICP-MS analysis. Samples from 2011 campaign have been measured for metal concentration following the same experimental procedure given in Arruti et al. (2011) to be included in the whole dataset used in this work. As a result, a total number of 138 samples were available for model development after an outlier removal process –89 from the period of 2008-2009 and 49 from 2011–, from which the following percentages were below the method detection limit: 53 % for As, 48 % for Cd, 23 % for Ni and a 6 % for Pb.

This work is based on the hypothesis that at a certain location there is an intrinsic relationship between metal(loid) immission concentrations, highly influenced by the anthropogenic activities, such as local industry, road traffic, or residential combustion; and the meteorological conditions and the presence of other major air pollutants in the ambient air, which is mainly due to local anthropogenic activities as well and would indirectly reflect the impact of anthropogenic emissions and the fate of pollutants once they are release to the atmosphere. Therefore, the input variables for statistical model development in this study consist of major air pollutant immission concentrations and meteorological variables, whose values are monitored in real time by local authorities through air quality networks and consequently are available at no additional cost. Major air pollutant data, namely concentrations of nitrogen oxides (NO_x), Sulphur dioxide (SO_2), tropospheric ozone (O_3) and PM_{10}, were obtained from the records of the air quality monitoring station located 1 km far from the sampling site, in Tetuán (43º 28’ 0” N, 3º 47’ 29” W at 24 m above sea level), which belongs to the Regional Air Quality Network (CIMA, 2015). Meteorological data, namely temperature (T), relative humidity (RH), wind direction (WD), wind speed (WS), pressure (P), solar radiation (SR) and cumulative precipitation (PP), and atmospheric soundings were provided by the monitoring station (CMT) of the State Meteorological Agency in Cueto (43º 29’ 28” N, 3º 48’ 2” W and at 52 m above sea level), at 2 km from the sampling site. Fig. 1 shows a map with the location of the sampling sites.

2.2. Importance of data splitting on statistical modelling

There are three main aspects that strongly influence the statistical model performance, especially artificial neural network models, and should be taken into account prior to the development of the models: (i) the quality of data; (ii) the number of data, and (iii) the data splitting. The quality of data used for the training directly affects the accuracy of the model. In addition, Flood et al. (1994) also noted that the performance of a model strongly depends on the number of available training samples: the more number of samples

available for model training/calibration, the more potentially accurate the model can be. Nevertheless, from a practical point of view, it is not always feasible having sufficient data for the training/calibration of a model, especially for certain applications such as the estimation of heavy metal levels in ambient air. Hence, the proportion of samples to include in each of the subsets becomes a very important issue.

Commonly, databases have to be divided into three different subsets to be used for statistical model development: training (calibration), test (cross-validation) and validation (external validation). As its name suggests, the former is used to adjust model parameters until an optimal set is found (train/calibrate the model), the following is used during training to avoid over-fitting, and the latter is employed to evaluate the model performance with respect to its capacity to provide estimations when using samples not previously considered in the training step, which is commonly referred to as testing the generalization ability of the model. The external validation of a model should be carried out by using an external validation subset of samples whose values are within the same range than those of the training subset. Otherwise, poor predictive ability is probable to be achieved, because normally a model is unable to make an extrapolation beyond the range of the training dataset, only interpolation.

Furthermore, a number of studies have reported that the data splitting step can decisively affect the model performance, especially artificial neural network models (Tokar and Johnson, 1999; Maier et al., 2000; Maier et al., 2010). Therefore, optimal data division into subsets is a crucial matter.

From a practical point of view, there are two different approaches of addressing this problem. On the one hand, approach A would correspond to use the data of previous years for model development, validating the model afterwards with the available data of the current year. This would be the real situation for authorities to use the developed model in the future. Authorities in practice are supposed to use the metal concentration data from precedent years along with major air pollutant concentration and meteorological information from the current year to develop models to estimate the pollutant concentrations for the current year. Afterwards, the resulting estimations should be compared with data from experimental measurements to test if those estimations are in fact acceptable. Whereas, on the other hand, approach B would consist of dividing the entire dataset into the aforementioned three sets disregarding the year to which they belong.

2.3. Description of methodology

Fig. 2 shows a flow diagram of the proposed methodology for the development of air quality models. It consists of five main steps: (i) removal of multivariate outliers, (ii) determination of samples that are above the method detection limit by means of a classification tree, (iii) data splitting into the required subsets for model calibration according to the two approaches described above, (iv) model development, and (v) model performance evaluation to check if air quality assessment is valid.

2.3.1. Multivariate outlier removal

Multivariate outlier identification and removal method was based on Mahalanobis distance. It is a well-known classical approach that computes the Mahalanobis distance (MD) of each observation as an indicative measure of the distance of each data point from the centre of the multivariate data cloud. By convention, this method identifies as outliers those observations with a large MD (exceeding the 99% quantile of a chi-square distribution).

2.3.2. Classification tree

The number of samples of metal concentration is generally limited because of the difficulty associated with their analytical determination. In addition, it may be possible that a great amount of these pollutant levels are below the method detection limit in those zones and agglomerations where objective estimation techniques are allowed to be used as air quality assessment tools. The common strategy for those samples below the method detection limit is to be quantified as half the method detection limit. Nevertheless, to conduct such substitution for a considerable number of samples especially in a database modest in size could make a negative impact on the subsequent model performance. Following the approach proposed by Zalel et al. (2015) to estimate airborne benzo(a)pyrene concentrations in the Czech Republic, a binary classification tree prior to model development has been applied to obtain, based on the observed independent and dependent variables, a set of rules that allow to estimate if the metal concentration of an specific sample is above or below a threshold value set as the method detection limit. The classification tree allow us identifying whether the combination of input variables corresponding to a certain sample represents conditions that are expected to be associated with a metal concentration above the method detection limit. In that case, the mentioned sample is included in the database that will be used afterwards for the quantitative estimation of the levels of that specific metal through model development. Otherwise, a preset estimation value equal to half the value of the method detection limit is assigned.

2.3.3. Data splitting

Given the importance of data splitting, highlighted in section 2.2, a number of studies have dealt with the problem of selecting representative subsets from a set of samples (Wu et al., 1996; Tominaga 1998; Sales et al., 2000; Bowden et al., 2002; Daszykowski et al., 2002; Wu et al., 2012). Random data splitting is a common method due to its inherent simplicity and because the data randomly extracted from a larger set follows the statistical distribution of the entire set, however neither does this method ensure the representatitviy of the extracted subsets nor does it ensure that the samples on the boundaries are included in the training/calibration subset, what would help to minimize possible extrapolation problems. Therefore, other techniques to extract data based on criteria established under more rigorous scientific basis are preferable. In this context, Kennard and Stone (1969) proposed an algorithm that measures the similarity between samples on the basis of the Euclidean distance. This method, as well as a group of other methods, e.g. DUPLEX algorithm, has been widely applied since then. In this context, Saptoro et al. (2012) proposed a modification of the Kennard-Stone algorithm, namely Kennard-Stone algorithm based on Mahalanobis distance (MDKS), to perform the data partition to develop artificial neural network models. Better performance results were found for this method when compared with the standard Kennard-Stone algorithm and other related algorithms such as the SPXY method, a data division method based on joint x-y distances and proposed by (Galvão et al., 2005). In this work the MDKS algorithm has been applied.

2.3.4. Model development

PLSR is a statistical method that as multiple linear regression (MLR) performs a linear combination of the predictors that best correlates the outcome. However, these predictors, as with principal component analysis (PCA), are actually components or *latent variables* created as a combination of the original input variables, but with the constraint of describing as much as possible the covariance between the independent and the dependent variables, whereas PCA focuses on the variance of the independent variables. The number of latent variables to be considered has been determined by a cross-validation. The PLSR models in this study have been developed using MATLAB (MathWorks, Inc.).

ANNs are computational systems that emulate the behavior of neural nervous systems of animals. They are composed by simple processing elements known as artificial neurons or units connected to one another on a weighted basis and organized in layers. The network architecture for the ANN models in this work corresponds to a Multilayer perceptron (MLP) that consisted of an input layer (each input unit represents one input variable), followed by a single hidden layer before an output layer (with just one

unit, containing the dependent variable or immission concentration of Pb, As, Cd or Ni). The ANNs have been trained using a Levenberg-Marquardt learning algorithm to update the set of connecting weights and with a sigmoid activation function in the hidden layer and a linear activation function in the output layer. Different number of units in the hidden layer have been tested and, finally, those that provided better results have been selected. The models have been developed using the Neural Network Toolbox for MATLAB (MathWorks, Inc.).

2.3.5. Performance criteria for model evaluation

Air quality model performance evaluation has become an important issue because the use of models for policy support has broaden. This is addressed in several documents published by policy-making authorities (ASTM standard D6589; 2005, EPA; 2009; Denby, 2010; Derwent et al., 2010). It is commonly conducted by means of a statistical performance analysis, which consists of comparing estimations against the corresponding observations using statistical indicators to evaluate the ability of the model to reproduce the measured concentrations. In this sense, Boylan and Russell (2006) define performance criteria as the minimum quality levels of the mentioned statistical parameters that a model has to comply with to be considered as acceptable for policy applications. These series of parameters allow making a general analysis of model performance. However, they provide insufficient insight on whether the models have satisfactorily reached an adequate quality level for a given application such as policy support (Boylan and Russel, 2006).

The quality level of the performance of objective estimation techniques has to be evaluated in the context of the European Air Quality Directives. The uncertainty for objective estimation is defined as the maximum deviation of the measured and calculated concentration levels over the period considered, by the limit value/target value, without taking into account the timing of the events. In this work, uncertainty requirements have been interpreted through the consideration of two indices: the relative maximum error without timing (RME), which is defined as the largest concentration difference of all percentile (p) differences between the observed concentration ($C_{O,p}$) and the estimated concentration ($C_{E,p}$) normalized by the respective measured value (Flemming and Stern, 2007; Borrego et al., 2008), as given by Eq. (1), and the relative directive error (RDE), the difference between the closest observed concentration to the limit/target value ($C_{O,LV}$) and the correspondingly ranked modelled concentration ($C_{E,LV}$) normalized by the limit/target value (LV/TV) (Denby, 2010), as calculated by Eq. (2).

$$RME = \max\left(\left|C_{O,p} - C_{E,p}\right|\right)/C_{O,p} \quad (1)$$

$$RDE = |C_{O,LV} - C_{E,LV}|/LV \quad (2)$$

According to Directive 2008/50/CE (EC, 2008), the data quality objective regarding objective estimation uncertainty is 100% and, therefore, that is the maximum value allowed for these indices.

Additionally, because limit/target values of metals in European Air Quality Directives are provided as annual mean values, the adequacy of mean value estimation is to be tested. This is done by means of the fractional bias:

$$FB = \frac{\overline{C_O} - \overline{C_E}}{0.5\,(\overline{C_O} + \overline{C_E})} \quad (3)$$

where $\overline{C_O}$ and $\overline{C_E}$ are the mean values (ng/m^3) of the observations and the estimations, respectively.

Two limit situations have been considered depending on if the estimated mean value is double or half the observed mean value, which corresponds to a FB index values equal to -2/3 and 2/3, respectively. Therefore, an FB index value outside the range of -2/3 and 2/3 will be considered unacceptable.

As RDE is mainly related to the exceedance of limit values, it provides only a partial view of the strengths and weaknesses of a given model application (Pederzoli et al., 2011). Thus, in order to deepen the analysis of model performance, a set of additional statistical indicators was required. In this work, some statistical parameters included in the EURASAP model validation kit were considered. This includes: the correlation coefficient (r), the root mean squared error (RMSE), the normalized mean squared error (NMSE) and the fractional variance (FV).

However, up until now, not many minimum quality criteria for these statistical indicators have been proposed in the literature regarding model performance evaluation and not always in a harmonized way. In this context, Kumar et al. (1993) proposed that the normalized mean squared error (NMSE) should be lower or equal to 0.5 and the fractional variance (FV) should be within the range from -0.5 to 0.5. These criteria have been applied in this work.

Additionally, in a more illustrative manner, a graphical approach has been applied to perform an easy comparison between model results and an evaluation with respect to their level of accuracy for policy applications. This representation, known as the target diagram, was proposed by Pederzoli et al. (2011) and has been developed as a modification of the Taylor diagram (Taylor, 2001) by the Joint Research Centre (JRC) of the European Commission within the framework of the Forum for Air Quality Modelling in Europe (FAIRMODE) and introduced inside the DELTA Tool (Thunis et al., 2012). It compares the mean bias, *y*-axis, as given by Eq. (4), and the centered mean squared error (CRMSE), *x*-axis, as given by Eq. (5), both normalized by the standard deviation of the observations (σ_O). In addition, it includes a

circumference of unit radius, which defines the acceptable limit value of model efficiency, MEF, (Eq. 6). According to this representation, the closest the model results are to the center of the circumference the higher the model performance. However, the model results must be inside the circumference to be considered acceptable, which means in practice that the normalized bias and CRMSE must be lower than 1 and a MEF must have a positive value.

$$\text{Bias}=\frac{1}{N}\sum_{i=1}^{N}\left(C_{E,i}-C_{O,i}\right)=\overline{C_E}-\overline{C_O} \quad (4)$$

$$\text{CRMSE}=\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left[\left(C_{E,i}-\overline{C_E}\right)-\left(C_{O,i}-\overline{C_O}\right)\right]^2} \quad (5)$$

where N is the total number of samples.

$$\text{MEF}=1-\left(\frac{\text{RMSE}}{\sigma_O}\right)^2 \quad (6)$$

2.4. Input variables related to atmospheric stability

Atmospheric stability plays an important role in air quality studies as it strongly affects the dispersion of pollutants in the atmosphere and it is indirectly introduced as an input in deterministic models. For these reasons, four meteorological variables regarded as good candidates to be introduced in the models as additional input variables to properly represent atmospheric stability have been considered: mixing height, the Richardson number, inversion and the Pasquill-Gifford atmospheric stability class. The value of these variables was determined from vertical temperature profiles of atmospheric soundings conducted by the Spanish State Meteorology Agency (AEMET) in Santander (Fig. 1).

The mixing height (MH) is the thickness of the air layer adjacent to the ground over which pollutants are mixed by convection or mechanical turbulence and, therefore, delimits the volume available for its dispersion (Beyrich, 1997; Seibert et al., 2000). In this study, the mixing height was determined applying the Holzworth method (Holzworth, 1972), a graphical approach of widespread use in the literature for this purpose. According to this method, the mixing height is given by the intersection between the dry adiabatic profile and the vertical temperature profile of an atmospheric sounding. In addition, those cases where inversion (I) occurred (i.e. temperature increased with height) were also identified (1: no inversion, 2: moderate inversion; 3: complete inversion).

Pasquill (1961) proposed a classification for the atmospheric stability in terms of increasing order from very unstable (A) to moderately stable (F). Thus, an atmosphere with an A stability class favors strong mixing whereas classes E or F impede the dispersion. In contrast, atmospheric stability can be numerically accounted for by means of the Richardson number (Ri), a dimensionless number that

determines stability through the relationship of convective and mechanical forces that take part in the atmospheric movements. Specifically, it is the ratio of the vertical temperature gradient to the squared vertical gradient of the wind speed. The numerator is related to the destabilizing forces that generate updrafts. The denominator is related to the kinetic energy that destroys updrafts (Woodward, 1998). Therefore, the larger the Richardson number, the more unstable is the atmosphere. In this study, the Richardson number was calculated in accordance to Woodward (1998) as follows:

$$Ri = \frac{g\,(T_t - T_b)\,(z_t - z_b)}{T_b\,(u_t - u_b)^2} \quad (7)$$

where g is the acceleration of gravity (m/s^2), T is the air temperature (K), z is sensor height (m) above ground level and u is the horizontal wind speed (m/s). The subscripts t and b are the top and bottom positions of vertical profiles, which in this work correspond to the mixing height and the ground level, respectively.

A number of equivalences between Pasquill-Gifford atmospheric stability classes and Ri values exist. In this work, the corresponding Pasquill-Gifford stability class (PG) was determined according to the equivalence proposed by Woodward (1998), and coded as 1-6, where 1 is very unstable (A class) and 6 is moderately stable (F class).

As two soundings a day (at 0h and 12h) were available in Santander station, two different values were considered for every input variable related to atmospheric stability: one in the morning and one in the evening, which are identified with the subscripts *am* and *pm*, respectively.

3. Results and discussion

3.1. Estimation of Pb, As, Cd and Ni immission levels in Santander

The air quality in relation to Pb, As, Cd and Ni has been assessed in Santander in this work by means of statistical models based on PLSR and ANNs following the methodology described in Fig. 2. The results of the best developed models considering approaches A and B (see section 2.2) are shown in Tables 1 and 2, respectively. Aside from external validation results, training results are also included for comparison purposes in order to test the generalization ability of the models.

Mathematically, every model performance indicator considered in this work has an ideal value of zero, except for the correlation coefficient and the model efficiency, whose ideal value is one. In this sense, except for Pb models, which exhibit a similar performance with both approaches, the external validation is in general better when considering approach B instead of approach A. This is probably due to the fact

that the models corresponding to Approach A have been trained with data from 2008 and 2009 and validated with data from 2011. Hence, there is a gap of one year between the data used to train/calibrate the models and the data used to test them and consequently there may have been a change in emission patterns especially if the period of time in which the study is focused is to be considered, because it corresponds to that of a great impact of the economic crisis on industrial activities that may have altered their production, with severe repercussions on emissions, in a local scale in particular. For that reason, it is assumed that according to approach A the model is not correctly calibrated to be able to predict the external validation dataset (2011) situation, with the consequent detriment in the generalization ability of the model. Therefore, the reason why approach B provides better results resides in the fact that the consideration of the three studied years simultaneously favors to some extent the attenuation of the differences between data relationships from 2011 and those of the preceding years. This is the main limitation of the proposed approach to be used as a routine air quality assessment tool by local authorities: its application is restricted to those cases in which the emission patterns are maintained over time, which is an acceptable assumption in the short-term but rarely valid in the long-term. For that reason, the models should be periodically updated with at least a few number of new measurements –control measurements– to guarantee that they are not out to date.

Regarding the evaluation of the presented approaches as tools for air quality assessment in the context of the European Directives, it would be strictly sufficient that the model estimations comply with the uncertainty and the mean value quality requirements because limit/objective values of Pb, As, Cd and Ni are given in mean concentrations over a calendar year. As seen in Table 2, all the best developed models considering approach B fulfill these requirements because both uncertainty indices, RME and RDE, are below 100% and the FB index is within the range of -2/3 and 2/3, which corresponds to half and double mean observed concentration, respectively. In fact, RDE values range up to 30%, with half of them below 10%. This means that, the difference between the observed value that is closest to the limit/target value and the respective estimation is lower than 30% of the actual limit/target value and, often, even lower than 10% of the limit/target value. Moreover, PLSR FB index values are in general below those of ANNs in both training and external validation steps.

Nevertheless, apart from that, local authorities would also be interested in being cognizant of the time variations of the studied pollutants or in being able to provide estimations with smaller time resolution different from an annual basis. Analyzing the values of the remaining statistical parameters, which in this

case give an idea on how well the models are able to estimate the observed concentrations of the individual 24-h samples, it is observed that there are general difficulties in providing accurate estimations of the 24-h mean concentrations. The NMSE and FV tend to be greater than 0.5. In addition, FV index values are all positive, which indicates that the estimated values show less variation than the observations. The estimations seem to vary around a value within the range of the dependent variable and close to the average, therefore, there is an overestimation of the lower values and more importantly an underestimation of the higher values. In addition, except for those of As, the correlation coefficient values are poor, lower than or equal to 0.5 in most cases. Nonetheless, there is no evidence of a superior performance of one technique over the other: PLSR and ANN models exhibit a comparable equivalent performance, what may be interpreted as that the relationship between model inputs and immission concentrations of Pb, As, Cd and Ni may tend to be linear rather than non-linear, at least as far as the available data are concerned. This idea is reinforced by the simplicity of most of the best neural network architectures developed, involving only a single hidden layer with just a few hidden neurons.

Fig. 3 represents the training and external validation results of the PLSR and ANN models developed for Pb, As, Cd and Ni taking into consideration approach B and the graphical approach described in section 2.3.5. Although, the CRMSE is always positive by its own mathematical definition (Eq. 5), a negative sign has been assigned to distinguish those situations when the standard deviation of the estimations is lower than the corresponding to the observations (Pederzoli et al., 2011). All the model results are located in the left side of the diagram, i.e. there are negative CRMSE values, indicating that the estimations vary within a narrower range than observations. This is also supported by the positive values of the FV index, as mentioned before (Table 2). Furthermore, as shown in Fig. 3, the external validation results for Pb and Cd models in approach B have a negative model efficiency and consequently they are out from the circumference. According to Joliff et al. (2009) the fact that the estimated values are inside the circumference, i.e. the model efficiency is greater than zero, means that the model is a better predictor of the observations than a constant value set to the average of observations.

3.2. Influence of input meteorological variables related with atmospheric stability

3.2.1. Classification of independent variables according to their correlation coefficient

The relative importance of every independent variable to be used as input variable to the model has been determined through the corresponding Pearson correlation coefficient with respect to the output variables, so that the more significant inputs can be identified. This coefficient evaluates the joint variation degree

or the similarity of two variables, quantifying the linear relationship existing between them. Tables 3 and 4 shows a classification of the independent variables – major pollutant concentrations and meteorological variables including those related to atmospheric stability – in descending order of their correlation coefficient in relation to each dependent variable, i.e. metal(loid) immission concentration. Additionally, the respective *p-value* is also presented. As two different approaches have been considered, approaches A and B (see section 2.2), Table 3 makes reference to the data of the period 2008-2009, whereas Table 4 presents results for the complete dataset, including 2011.

First of all, there is a non-strong correlation between the independent variables and the dependent variables individually. However, the consideration of the entire set of independent variables altogether may change the individual influence due to synergic effects.

Pb and Ni seem to be more related to major air pollutant concentration, especially PM_{10} concentrations. However, this tendency is not observed for As and Cd.

The consideration of samples from 2011 leads to a significant modification in the correlation coefficients and affects the relative position of each variable in the classification. As mentioned before in section 3.1, this may be caused by a change in the emission patterns in the Region, probably associated with the fluctuations in the production of industrial activities as a consequence of the economic and financial crisis. Not in vain, the period of study extends from 2008 to 2011, which is the period of time with more detrimental effects for the economy.

However, in any case, some of the meteorological variables related to atmospheric stability are ranked in the top positions for every pollutant. What is more, for every pollutant in the period 2008-2011, at least one of the four considered new input variables was among those statistically significant with a 90% confidence level. In general terms, based on the ranking positions of these new variables, the Pasquill-Gifford stability class appears to be the most significant variable.

3.2.2. Estimation of Pb, Cd and Ni immission levels with the new meteorological variables

In this study, it is proposed to consider meteorological variables related with atmospheric stability as additional input variables to the models to further improve their performance. As a consequence, PLSR and ANN models have been developed considering as supplementary input meteorological variables those that, based on the individual correlation coefficient (Table 4), presented most relevance for each pollutant. Table 5 shows the results for Pb and Ni. A reduction in the RME and RDE values together with an increment of the correlation coefficient values is observed when considering the morning Pasquill-

Gifford stability class (PG_{am}). Except for the Ni PLSR model, there is also an enhancement of the FB values of the external validation subset. The results for Cd are not presented because there was no significant improvement of the model performance. Fig. 4 and Fig. 5 show a comparison between the target diagram obtained for regular Pb and Ni models presented in section 3.1. and those obtained when considering the PG_{am} as an input. In general, there is a slight improvement of model results. In particular, it is observed that the consideration of the morning Pasquill-Gifford stability class as an additional input meteorological variable has a greater influence over the ANN model outcomes. It results in a decrease in the bias and the CRMSE values of the ANN external validation results for Pb what lead to a change in the MEF sign, which turn into a positive value positioned inside the circumference. Regarding Pb PLSR model results, they experienced a minor upgrade, moving slightly closer to the center.

With respect to Ni, the results of ANN models exhibit a more noticeable improvement than PLSR models, especially in terms of mean concentration estimation of the external validation subset, with a considerably reduced bias value. With respect to Ni PLSR results, although they still are inside the circumference and therefore they can still be considered valid, they experienced a slight deterioration regarding the external validation results.

Due to the considerable effort required for the user, as it entails the availability of atmospheric sounding data and an advanced data processing, this approach should only be taken into consideration in those situations in which there is room for improvement, i.e. an unsatisfactory model performance is obtained because the values of statistical parameters used to test it do not meet quality objectives.

4. Conclusions

In this study a proposal for a methodology to use statistical models based on PLSR and ANNs for air quality assessment in relation to EU regulated metal(loids) is presented. Two different approaches have been considered: (i) approach A, which would represent the actual situation for authorities regarding air quality assessment, and (ii) approach B, which would consist of using all the available information disregarding the period of time to which it belongs. These approaches have been tested using a dataset of Pb, As, Cd and Ni levels in Santander (Spain) in 2008, 2009 and 2011.

It is observed that statistical parameters used as indicators of model performance are in general closer to their ideal value when using approach B. This may be caused by a modification of the relationship between inputs and outputs of the models as a consequence of a change of emission patterns over time,

because they have been trained/calibrated with data from 2008-2009 and validated with data from 2011. Taking into account this consideration, in order to use approach A with confidence as a routine air quality assessment tool in the future, it would be indispensable to carry out brief control sampling campaigns to include a few new additional samples to keep the models up to date.

In the context of the European Framework Air Quality Directive, approach B would be considered a satisfactory air quality assessment tool in relation to Pb, As, Ni and Cd as it meets the quality requirements for estimations regarding uncertainty and accuracy in the mean value estimation. However, the models fail to estimate with precision daily concentrations. For that reason, the consideration of new input meteorological variables related to atmospheric stability to improve model performance has been tested with those variables that have shown a greater correlation with the levels of the EU regulated metals and metalloids in Santander. The consideration of the morning Pasquill-Gifford stability class as an additional meteorological input has yield a slight improvement in the model performance for Pb and Ni.

On the other hand, the similarity between results obtained applying the two different techniques considered, PLSR and ANN, indicates that at least for the available data the relationship between inputs and outputs may have a linear character. This is something to take into account for future model development, given the relative simplicity of linear techniques in comparison to non-linear techniques.

Acknowledgements

This work was supported by the Spanish Ministry of Economy and Competitiveness (MINECO) through the Projects CTM2010-16068 and CTM2013-43904R. Germán Santos thanks MINECO for his FPI research fellowship (BES-2011-047908). The authors also acknowledge the Spanish State Meteorology Agency (AEMET) for providing meteorological and atmospheric sounding data for the period of study.

References

1. Arruti, A., Fernández-Olmo, I., Irabien, A., 2011. Assessment of regional metal levels in ambient air by statistical regression models. J. Environ. Monit. 13(7), 1991-2000.
2. ASTM standard D6589, 2005. Standard Guide for Statistical Evaluation of Atmospheric Dispersion Model Performance (No. D6589). ASTM International, West Conshohocken, PA.

3. Beyrich, F., 1997. Mixing height estimation from sodar data - a critical discussion. Atmos. Environ. 31(23), 3941-3953.
4. Borrego, C., Monteiro, A., Ferreira, J., Miranda, A.I., Costa, A.M., Carvalho, A.C., Lopes, M., 2008. Procedures for estimation of modelling uncertainty in air quality assessment. Environ. Int. 34, 613-620.
5. Bowden, G.J., Maier, H.R., Dandy, G.C., 2002. Optimal division of data for neural network models in water resources applications. Water Resour. Res. 38(2), 1-11.
6. Boylan, J.W., Russell, A.G., 2006. PM and light extinction model performance metrics, goals, and criteria for three-dimensional air quality models. Atmos. Environ. 40, 4946-4959.
7. Centro de Investigación del Medio Ambiente (CIMA), 2015. Datos on-line de la Red de Control y vigilancia de la Calidad del Aire de Cantabria. Consejería de Medio Ambiente, Ordenación del Territorio y urbanismo del Gobierno de Cantabria. http://www.medioambientecantabria.es/calidad_aire/
8. Chelani, A.B., Gajchate, S.M., Tamhane, M., Hasan, M.Z., 2001. Statistical modelling of ambient air pollutants in Delhi. Water Air Soil Poll. 132, 315-331.
9. Chelani, A.B., Gajghate, D.G., Hasan, M.Z., 2002. Prediction of ambient PM_{10} and toxic metals using artificial neural networks. J. Air Waste Manage. Assoc. 52(7), 805-810.
10. Daly, A., Zannetti, P., 2007. Air Pollution Modeling – an overview. In: Zannetti, P., Al-Ajmi, D., Al-Rashied, S. (Eds.) Ambient air pollution, Arab School for Science and Technology (ASST) and The EnviroComp Institute.
11. Daszykowski, M., Walczak, B., Massart, D.L., 2002. Representative subset selection. Anal. Chim. Acta 468, 91-103.
12. Denby, B., 2010. Guidance on the use of the models for the European Air Quality Directive. A Working document of the Forum for Air Quality Modelling in Europe, FAIRMODE. ETC/ACC Report.
13. Derwent, D., Fraser, A., Abott, J., Willis, P., Murrells, T., 2010. Evaluating the performance of air quality models (No. Issue 3). Department for Environment and Rural Affairs.
14. EPA, 2009. Guidance document on the Development, Evaluation, and Application of Regulatory Environmental Models. No. EPA/100/K-09/003. U.S. Environmental Protection Agency.

15. European Commission (EC), 2004. Council Directive 2004/107/EC, Directive of the European Parliament and of the Council of 15 December 2004 relating to arsenic, cadmium, mercury, nickel and polycyclic aromatic hydrocarbons in ambient air. The European Parliament and the Council of the European Union. Off. J. L23, 3-16.
16. European Commission (EC), 2008. Council Directive 2008/50/EC, Directive of the European Parliament and of the Council of 21 May 2008 on ambient air quality and cleaner air for Europe. The European Parliament and the Council of the European Union. Off. J. L152, 1-44.
17. Flemming, J., Stern, R., 2007. Testing model accuracy measures according to the EU directives – examples using the chemical transport model REM-CALGRID. Atmos. Environ. 41, 9206-9216.
18. Flood, I., Kartam, N., 1994. Neural networks in civil engineering. I: Principles and understanding. J. Comput. Civil Eng. 8(2), 131-148.
19. Galvão, R.K.H., Araujo, M.C.U., José, G.E., Pontes, M.J.C., Silva, E.C., Saldanha, E.C.B., 2005. A method for calibration and validation subset partitioning. Talanta 67, 736-740.
20. Hernández, E., Martín, F., Valero, F., 1992. Statistical forecast models for daily air particulate iron and lead concentrations for Madrid, Spain. Atmos. Environ. 26B, 107-116.
21. Holzworth, G.C., 1972. Mixing heights, wind speeds and potential for urban pollution throughout the contiguous United States. Office al Air Programs Ap-101, U.S.EPA (NTIS PB 207 103).
22. Instituto Nacional de Estadística (INE), 2014. INEbase. Cifras oficiales de población de los municipios españoles. http://www.ine.es/dynt3/inebase/es/index.html?padre=517&dh=1
23. Jolliff, J.K., Stow, C.A., McGillicuddy, D.J., Allen, J.I., Rose, K.A., Wallhead, P., 2009. Skill assessment for coupled biological/physical models of marine systems. J. Marine Syst. 76, 4-15.
24. Kennard, R.W., Stone, L.A., 1969. Computer aided design of experiments. Technometrics, 11(1), 137-148.
25. Kumar, A., Luo, J., Bennett, G., 1993. Statistical evaluation of lower flammability distance (LFD) using four hazardous release models. Process. Saf. Prog. 12, 1-11.
26. Li, Y., Luo, F., Jiang, Y., Lu, Y., Huang, J., Zhang, Z.., 2009. A prediction model of occupational manganese exposure based on artificial neural network. Toxicol. Mech. Method. 19(5), 337-345.

27. Maier, H.R., Dandy, G.C., 2000. Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications. Environ. Modell. Softw. 15, 101-124.

28. Maier, H.R., Jain, A., Dandy, G.C., Sudheer, K.P., 2010. Methods used for the development of neural networks for the prediction of water resource variables in river systems: current status and future directions. Environ. Modell. Softw. 25, 891-909.

29. Pasquill, F., 1961. The estimation for the dispersion of wind borne material. Meteorol. Mag. 90:34-49.

30. Pederzoli, A., Thunis, P., Georgieva, E., Borge, R., Carruthers, D., 2011. Performance criteria for the benchmarking of air quality model regulatory applications: the "target" approach. 14th Conference on harmonisation within atmospheric dispersion modelling for regulatory purposes, 297-301.

31. Sales, F., Rius, A., Callao, M.P., Rius, F.X., 2000. Standarization of a multivariate calibration model applied to the determination of chromium in tanning sewage. Talanta 52, 329-336.

32. Saptoro, A., Tadé, M.O., Vuthaluru, H., 2012. A modified Kennard-Stone algorithm for optimal division of data for developing artificial neural network models. Chem. Prod. Process Model. 7(1), 13.

33. Seibert, P., Beyrich, F., Gryning, S.E., Joffre, S., Rasmussen, A., Tercier, P., 2000. Review and intercomparison of operational methods for the determination of the mixing height. Atmos. Environ. 34, 1001-1027.

34. Stow, C.A., Jolliff, J., McGillicuddy, D.J., Doney, S.C., Allen, J.I., Friedrichs, M.A.M., Rose, K.A., Wallhead, P., 2009. Skill assessment for coupled biological/physical models of marine systems. J. Mar. Syst. 76, 4-15.

35. Taylor, K.E., 2001. Summarizing multiple aspects of model performance in a single diagram. J. Geophys. Res. 106, No. D7, 7183-7192.

36. Thunis, P., Georgieva, E., Pederzoli, A., 2012. A tool to evaluate air quality model performances in regulatory applications. Environ. Modell. Softw. 38, 220-230.

37. Tokar, A.S., Johnson, P.A., 1999. Rain fall-runoff modeling using artificial neural networks. J. Hydrol. Eng. 4(3), 232-239.

38. Tominaga, Y., 1998. Representative subset selection using genetic algorithms. Chemometr. Intell. Lab. syst. 43, 157-163.

39. Vicente, A.B., Jordán, M.M., Sanfeliu, T., Sánchez, A., Esteban, M.D., 2012. Air pollution prediction models of particles, As, Cd, Ni and Pb in a highly industrialized area in Castellón (NE, Spain*).* Environ. Earth Sci., 66(3), 879-888.

40. Woodward, J.L., 1998. Improving the effect of atmospheric stability class for dispersion modeling. Process. Saf. Prog. 17(1), 1-8.

41. Wu, W., May, R., Dandy, G.C., Maier, H.R., 2012. A method for comparing data splitting approaches for developing hydrological ANN models. Managing resources of a limited planet: proceedings of the 6th International congress on environmental modelling and software (iEMSs 2012), 1620-1627.

42. Wu, W., Walczak, B., Massart, D.L., Heuerding, S., Erni, F., Last, I.R., Prebble, K.A., 1996. Artificial neural networks in classification of NIR spectral data: Design of the training set. Chemometr. Intell. Lab. Syst. 33, 35-46.

43. Zalel, A., Yuval, Svecova, V., Sram, R.J., Bartonova, A., Broday, D.M., 2015. Modelling airborne benzo(a)pyrene concentrations in the Czech Republic. Atmos. Environ. 101, 166-176.

Figure captions

Figure 1: Location of the sampling site and the air quality and meteorological monitoring stations.

Figure 2: Flow diagram of the proposed methodology for objective estimation air quality assessment.

Figure 3: Target diagram of PLSR and ANN model results for Pb (red), As (blue), Ni (green) and Cd (yellow). Data points related to PLSR are represented by a circle and those related to ANNs are represented by a rhombus. Filled and hollow markers have been used to depict training (T) and external validation (EV) results, respectively.

Figure 4: Target diagram of PLSR and ANN model results for Pb to determine how considering PG_{am} as an additional input affects their performance.

Figure 5: Target diagram of PLSR and ANN model results for Ni to determine how considering PG_{am} as an additional input affects their performance.

Table 1

Click here to download Table: Table 1.docx

Table 1. PLSR and ANN model results considering approach A

Pollutant	Model	Subset[a]	RME (%)	RDE (%)	$\overline{C_O}$ (ng/m^3)	$\overline{C_E}$ (ng/m^3)	FB	r	RMSE (ng/m^3)	NMSE	FV
Pb	PLSR	T	50.9	5.0	6.7	6.7	0.000	0.424	6.77	1.02	0.73
		EV	69.5	8.6	13.2	9.1	0.370	0.304	13.32	1.48	1.16
	ANN	T	55.3	3.6	6.7	6.7	-0.005	0.636	5.77	0.74	0.46
		EV	74.6	8.2	13.2	8.7	0.407	0.404	12.98	1.46	1.00
As	PLSR	T	2.5	0.1	1.1	1.1	0.000	1.000	0.02	0.00	0.00
		EV	61.4	71.1	2.7	1.3	0.733	0.367	1.96	1.10	0.49
	ANN	T	35.6	14.2	1.1	1.0	0.076	0.801	0.50	0.23	1.07
		EV	79.8	96.2	2.7	0.1	0.971	0.258	2.22	1.90	1.55
Ni	PLSR	T	46.9	15.7	1.1	1.1	0.000	0.553	0.82	0.51	0.51
		EV	54.4	28.2	1.8	1.4	0.226	0.280	1.51	0.92	0.71
	ANN	T	58.7	12.4	0.9	1.3	-0.295	0.584	0.80	0.55	1.28
		EV	74.4	28.2	1.8	1.2	0.341	0.294	1.54	1.08	1.57
Cd	PLSR	T	70.0	24.0	0.3	0.3	0.000	0.354	0.29	0.93	0.81
		EV	60.7	5.0	0.2	0.4	-0.538	-0.071	0.30	1.14	0.52
	ANN	T	n.c.	n.c.	n.c.	n.c.	n.c.	n.c.	n.c.	n.c.	n.c.
		EV	n.c.	n.c.	n.c.	n.c.	n.c.	n.c.	n.c.	n.c.	n.c.

[a]T: Training; EV: External validation; n.c.: not calculated

Table 2

Click here to download Table: Table 2.docx

Table 2. PLSR and ANN model results for Pb, As, Cd and Ni considering approach B

Pollutant	Model	Subset[a]	RME (%)	RDE (%)	$\overline{C_O}$ (ng/m^3)	$\overline{C_E}$ (ng/m^3)	FB	r	RMSE (ng/m^3)	NMSE	FV
Pb	PLSR	T	62.0	7.58	7.6	7.6	0.000	0.366	9.4	1.55	0.71
		EV	60.7	8.15	10.8	9.9	0.090	0.257	10.4	1.01	0.77
	ANN	T	80.5	8.60	7.9	6.5	0.185	0.523	9.2	1.64	1.03
		EV	78.1	7.42	10.8	6.0	0.579	0.338	11.1	1.92	1.03
As	PLSR	T	13.0	2.83	0.7	0.7	0.000	0.971	0.1	0.04	0.12
		EV	47.8	16.01	1.0	0.8	0.268	0.714	0.6	0.42	0.43
	ANN	T	62.3	20.55	0.7	0.6	0.164	0.878	0.5	0.56	1.33
		EV	59.2	14.82	1.0	0.6	0.461	0.996	0.7	0.82	1.42
Ni	PLSR	T	45.2	13.75	1.1	1.1	0.000	0.488	0.8	0.58	0.36
		EV	32.5	9.19	1.5	1.5	0.008	0.530	0.9	0.38	0.46
	ANN	T	47.5	12.74	1.1	1.1	-0.010	0.564	0.7	0.44	0.68
		EV	51.8	14.11	1.5	1.1	0.316	0.517	1.0	0.63	0.79
Cd	PLSR	T	65.9	22.09	0.3	0.3	0.000	0.199	0.3	1.22	0.91
		EV	43.3	8.06	0.2	0.3	-0.492	0.160	0.2	0.97	1.38
	ANN	T	56.8	29.03	0.3	0.2	0.269	0.467	0.3	1.77	0.77
		EV	24.5	10.85	0.2	0.2	-0.182	-0.186	0.3	2.04	0.20

[a]T: Training; EV: External validation

Table 3

Click here to download Table: Table 3.docx

Table 3. Classification of independent variables according to correlation coefficient for the period 2008-2009

Ranking	Pb		As		Ni		Cd	
Position	variable	p-value	variable	p-value	variable	p-value	variable	p-value
1	NO_x	**0.003**	**Ri_{am}**	**0.002**	Ln PM_{10}	**0.001**	**MH_{pm}**	**0.014**
2	Ln PM_{10}	**0.010**	SD	**0.003**	**I_{pm}**	**0.002**	PP	0.194
3	O_3	**0.027**	T	**0.053**	RH	**0.012**	SE	0.251
4	**I_{pm}**	**0.031**	Ln PM_{10}	**0.081**	**PG_{am}**	**0.022**	**I_{am}**	0.314
5	**MH_{am}**	0.291	RS	0.107	**I_M**	**0.025**	O_3	0.366
6	P	0.307	WE	0.134	SO_2	**0.093**	**I_{pm}**	0.392
7	WE	0.361	SE	0.174	P	0.102	NO_x	0.422
8	RS	0.365	**PG_{pm}**	0.190	**MH_{am}**	0.109	P	0.481
9	T	0.419	PP	0.279	WS	0.141	T	0.533
10	**MH_{pm}**	0.467	**I_V**	0.310	**PG_{pm}**	0.161	WS	0.539
11	SO_2	0.553	O_3	0.432	NO_x	0.181	**PG_{pm}**	0.549
12	SD	0.555	**Ri_{pm}**	0.437	T	0.196	WE	0.553
13	PP	0.612	**MH_{pm}**	0.446	**MH_{pm}**	0.216	**MH_{am}**	0.577
14	**Ri_{pm}**	0.642	NO_x	0.446	SD	0.235	RS	0.654
15	**Ri_{am}**	0.705	P	0.526	RS	0.236	Ln PM_{10}	0.675
16	WD	0.710	**MH_{am}**	0.574	WE	0.391	**PG_{am}**	0.681
17	SE	0.730	RH	0.675	**Ri_{pm}**	0.793	SO_2	0.726
18	**PG_{pm}**	0.738	WD	0.679	SE	0.798	RH	0.822
19	**I_{am}**	0.861	**PG_{am}**	0.722	O_3	0.807	**Ri_{pm}**	0.832
20	WS	0.861	SO_2	0.790	**Ri_{am}**	0.883	SD	0.857
21	RH	0.886	**I_{am}**	0.809	WD	0.914	WD	0.871
22	**PG_{am}**	0.999	WS	0.886	PP	0.988	**Ri_{am}**	0.955

The meteorological variables related to atmospheric stability and the *p-values* of those variables statistically significant with a 90% confident level are pesented in bold.

Table 4

Click here to download Table: Table 4.docx

Table 4. Classification of independent variables according to correlation coefficient for the period 2008-2009 and 2011

Ranking	Pb		As		Ni		Cd	
Position	variable	p-value	variable	p-value	variable	p-value	variable	p-value
1	O_3	**0.001**	WS	**0.000**	WS	**0.000**	**PG_{pm}**	**0.005**
2	NO_x	**0.004**	WD	**0.000**	Ln PM_{10}	**0.001**	**MH_{pm}**	**0.005**
3	WD	**0.014**	SO_2	**0.000**	T	**0.015**	**PG_{am}**	**0.016**
4	**PG_{am}**	**0.049**	T	**0.000**	WD	**0.018**	SE	**0.021**
5	RH	**0.052**	NO_x	**0.000**	RH	**0.022**	**Ri_{pm}**	0.121
6	SE	**0.064**	SE	**0.000**	**PG_{am}**	**0.031**	RH	0.123
7	**MH_{am}**	**0.085**	**I_{am}**	**0.001**	**MH_{am}**	**0.041**	**I_{am}**	0.125
8	**PG_{pm}**	**0.088**	O_3	**0.046**	**I_{pm}**	**0.043**	WD	0.163
9	**I_{am}**	**0.097**	**I_{pm}**	**0.076**	**PG_{pm}**	**0.097**	P	0.182
10	Ln PM_{10}	0.130	Ln PM_{10}	0.100	SE	0.105	PP	0.212
11	**I_{pm}**	0.131	**Ri_{pm}**	0.113	RS	0.159	O_3	0.226
12	P	0.138	**PG_{pm}**	0.154	P	0.189	**MH_{am}**	0.247
13	SO_2	0.181	**PG_{am}**	0.170	**Ri_{pm}**	0.269	SD	0.261
14	T	0.208	**MH_{am}**	0.178	O_3	0.293	RS	0.293
15	**MH_{pm}**	0.239	WE	0.315	WE	0.323	WE	0.385
16	**Ri_{pm}**	0.242	RH	0.360	SO_2	0.361	**Ri_{am}**	0.386
17	**Ri_{am}**	0.283	PP	0.371	**Ri_{am}**	0.364	T	0.399
18	PP	0.341	**MH_{pm}**	0.408	**I_{am}**	0.387	WS	0.447
19	SD	0.344	RS	0.409	NO_x	0.432	**I_{pm}**	0.477
20	WS	0.352	SD	0.441	PP	0.436	SO_2	0.480
21	WE	0.425	**Ri_{am}**	0.459	SD	0.445	Ln PM_{10}	0.486
22	RS	0.483	P	0.480	**MH_{pm}**	0.486	NO_x	0.494

The meteorological variables related to atmospheric stability and the *p-values* of those variables statistically significant with a 90% confident level are pesented in bold.

Table 5

Click here to download Table: Table 5.docx

Table 5. PLSR and ANN model results for Pb and Ni with and without considering PG_{am} as input variable under approach B

Pollutant	Model	Subset[a]	RME (%)	RDE (%)	$\overline{C_O}$ (ng/m^3)	$\overline{C_E}$ (ng/m^3)	FB	r	RMSE (ng/m^3)	NMSE	FV
Pb	PLSR	T	62.0	7.58	7.6	7.6	0.000	0.366	9.4	1.55	0.71
		EV	60.7	8.15	10.8	9.9	0.090	0.257	10.4	1.01	0.77
	PLSR - PG_{am}	T	57.3	6.70	7.6	7.6	0.000	0.574	8.0	1.11	0.56
		EV	54.8	5.33	9.8	9.6	0.012	0.463	8.2	0.72	0.69
	ANN	T	80.5	8.60	7.9	6.5	0.185	0.523	9.2	1.64	1.03
		EV	78.1	7.42	10.8	6.0	0.579	0.338	11.1	1.92	1.03
	ANN - PG_{am}	T	57.3	6.49	7.4	9.5	-0.247	0.642	7.6	0.82	0.71
		EV	45.9	5.97	9.8	9.4	0.041	0.515	8.0	0.71	0.84
Ni	PLSR	T	45.2	13.75	1.1	1.1	0.000	0.488	0.8	0.58	0.36
		EV	32.5	9.19	1.5	1.5	0.008	0.530	0.9	0.38	0.46
	PLSR - PG_{am}	T	27.5	12.31	1.2	1.2	0.000	0.593	0.7	0.42	0.43
		EV	41.6	21.36	1.5	1.4	0.092	0.410	1.2	0.68	0.52
	ANN	T	47.5	12.74	1.1	1.1	-0.010	0.564	0.7	0.44	0.68
		EV	51.8	14.11	1.5	1.1	0.316	0.517	1.0	0.63	0.79
	ANN - PG_{am}	T	5.6	1.05	1.2	1.2	-0.025	0.958	0.2	0.05	0.06
		EV	29.7	14.98	1.5	1.4	0.103	0.479	1.2	0.70	0.23

[a]T: Training; EV: External validation

Figure 1

Click here to download high resolution image

Figure 2

Click here to download high resolution image

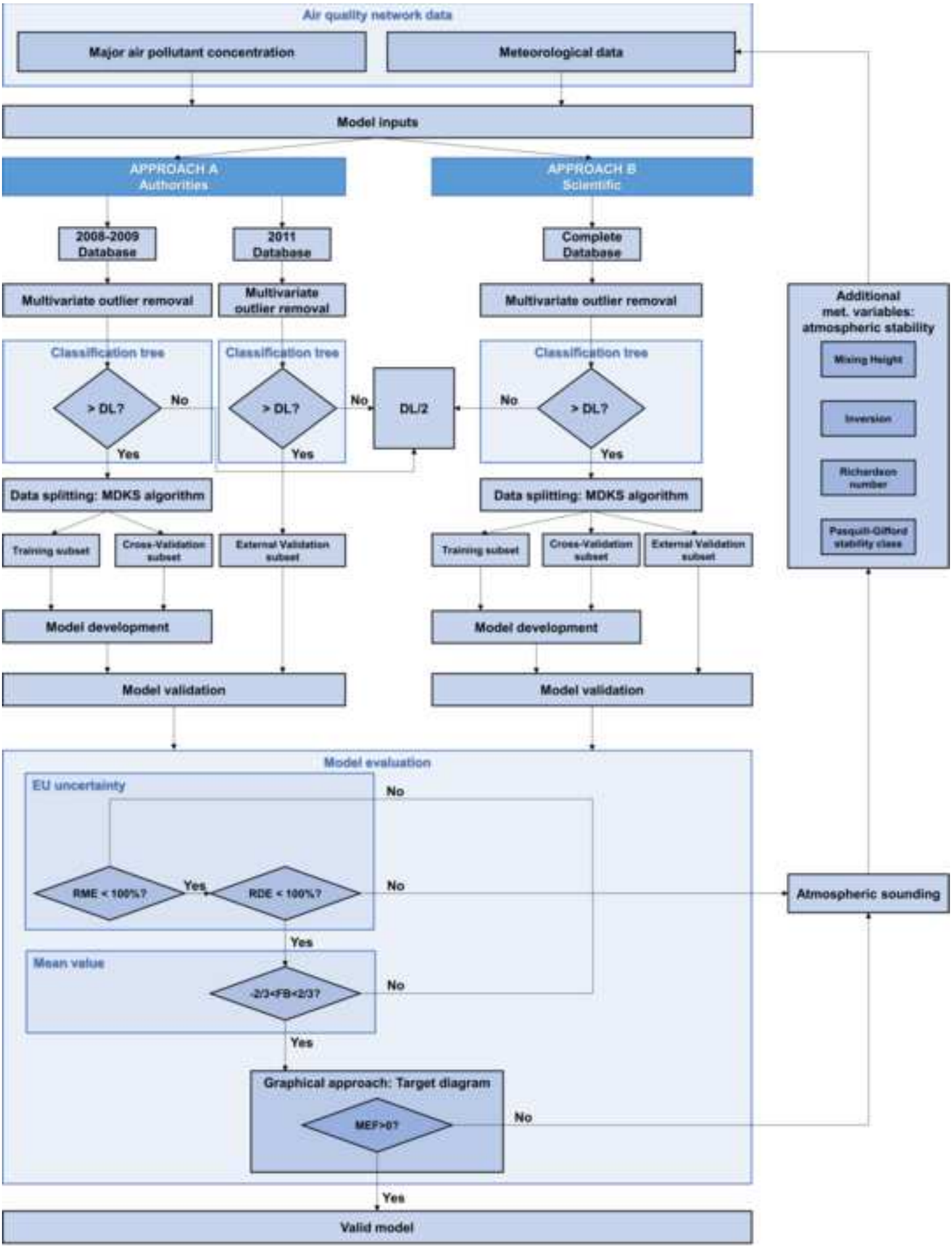

Figure 3

Click here to download high resolution image

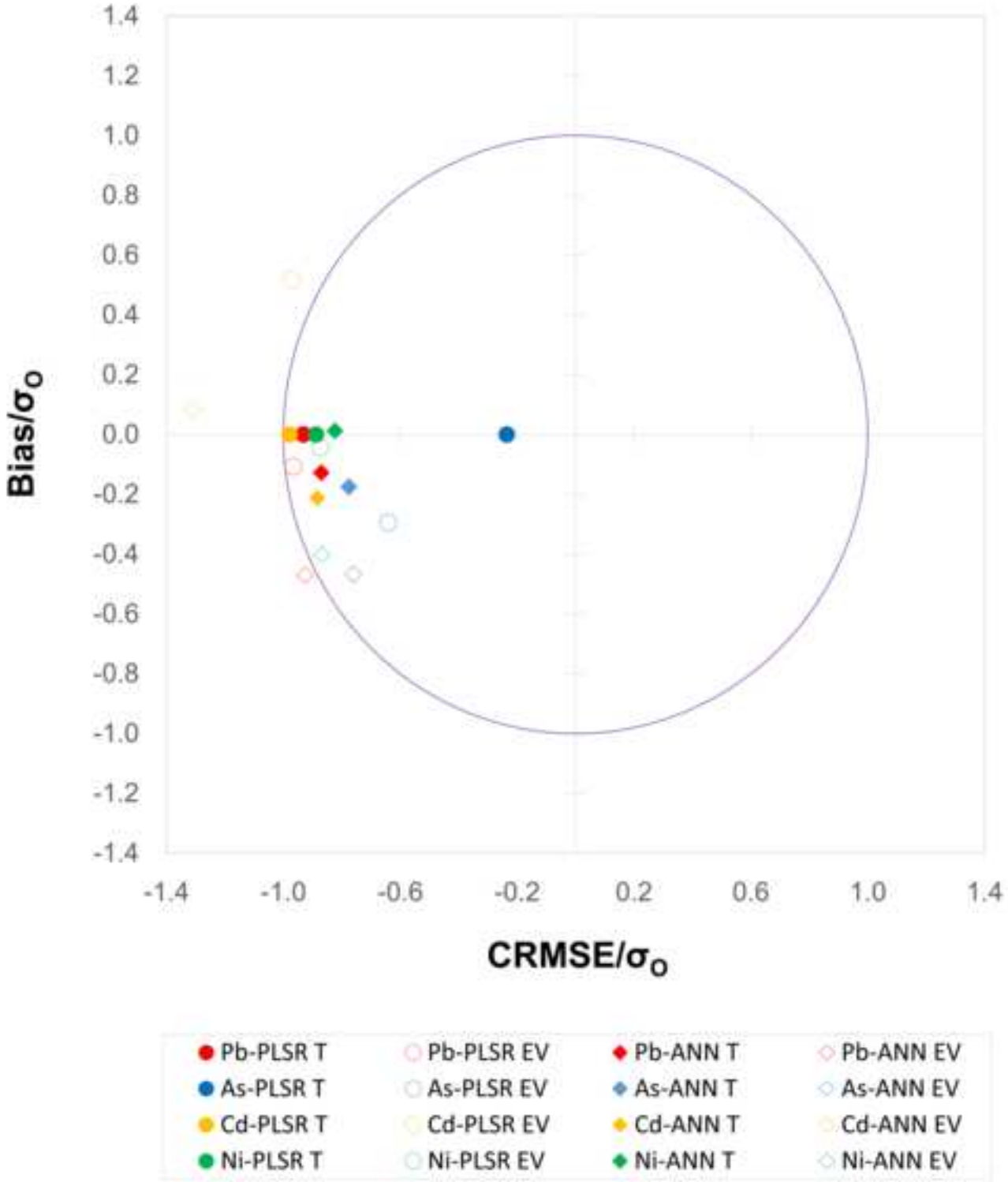

Figure 4

Click here to download high resolution image

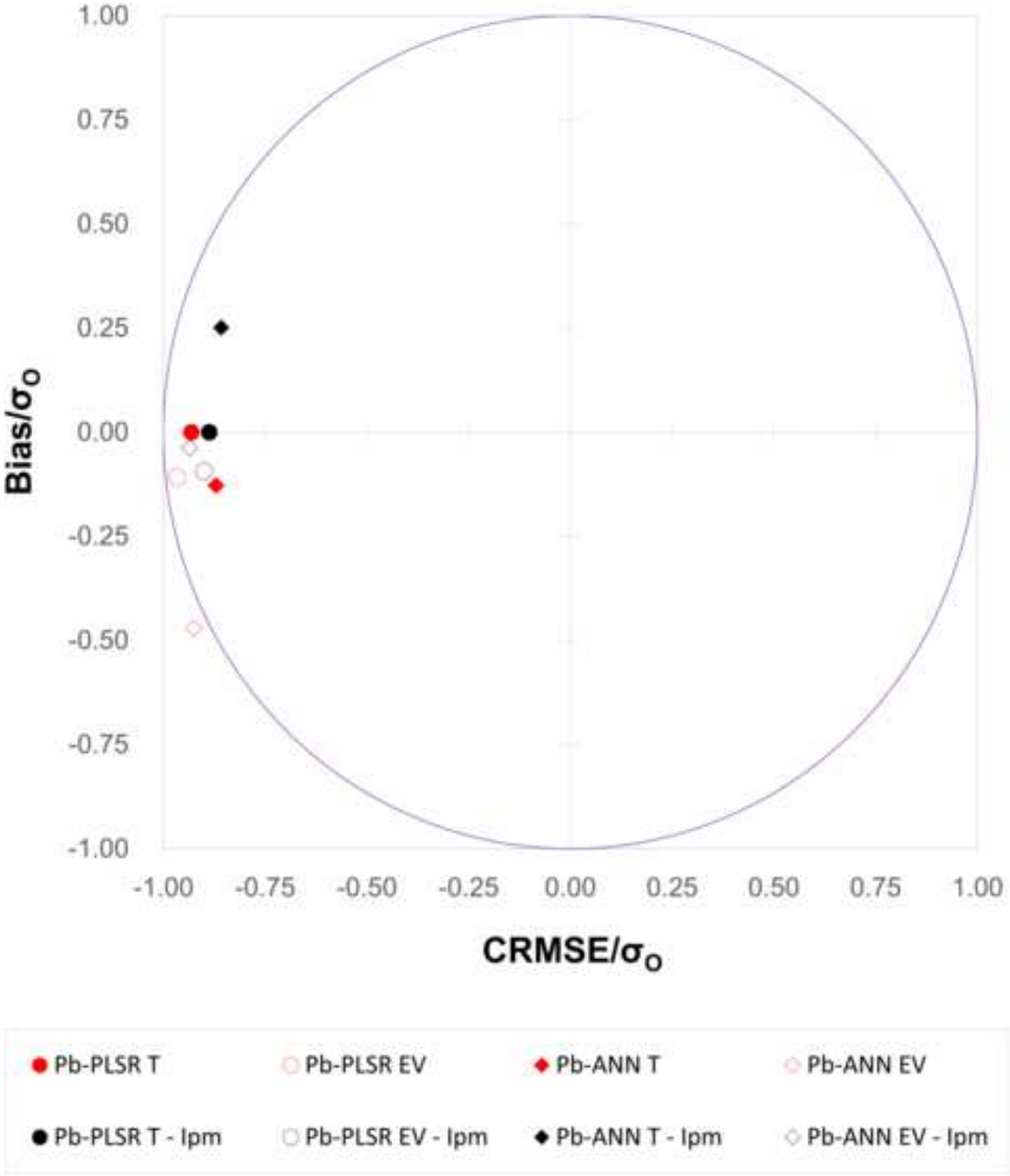

Figure 5

Click here to download high resolution image